A global analysis of the relationship between violence and life expectancy

Big Data in Social Sciences

Riccardo Omenti

University of Bologna

Introduction

Measuring **violence** experienced by a country is a challenging task.

- **Violence** is a multifaceted concept
- It depends on many different factors
- Multiple sources are needed to capture its complexities

Research Question

- The level of violence may affect the mortality of a country
- Today, we will harness two major data sources to explore the bivariate relationship between violence and mortality via:
  - Data on Violece from the Institute for Economics & Peace
  - Data on Mortality from the United Nations
- We will produce multiple descriptive plots with **ggplot**

ggplot

ggplot is an R package for producing statistical, or data, graphics Already available in

tidyverse Major advantages:

- creating graphs by combining independent components
- detailed theming system to generate nice-looking graphs
- intuitive grammar

Indicators for violence and mortality

- **Violence** $\rightarrow$ **Global Peaceful Index** (GPI) that measures the violence of a country across three dimensions:
  - ongoing domestic and international conflict
  - societal safety and security
  - militarization
- **Mortality** $\rightarrow$ **Life Expectancy at Birth** (e_0)
- Two *.txt* data files: *gpi_data.txt* and *life_exp_data.txt*

Upload data in Rstudio

Upload *tidyverse* and the data sets

```
#install.packages('ggthemes')
#install.packages('ggthemes')
library('RColorBrewer') # various qualitative color palettes
library('ggthemes') # niche themes in ggplot
library("tidyverse") # ggplot2
gpi_data <- read.table('Data/gpi_data.txt',header=T)
life_exp_data <- read.table('Data/life_exp_data.txt',header=T)
```

Link the two data sets by iso3 and year

```
data <- left_join(gpi_data,life_exp_data,by=c('iso3','Year'))
```

Question (1)

Create a plot displaying the evolution of violence over the time period 2008-2022 by world region

1. Calculate region-specific levels of violence

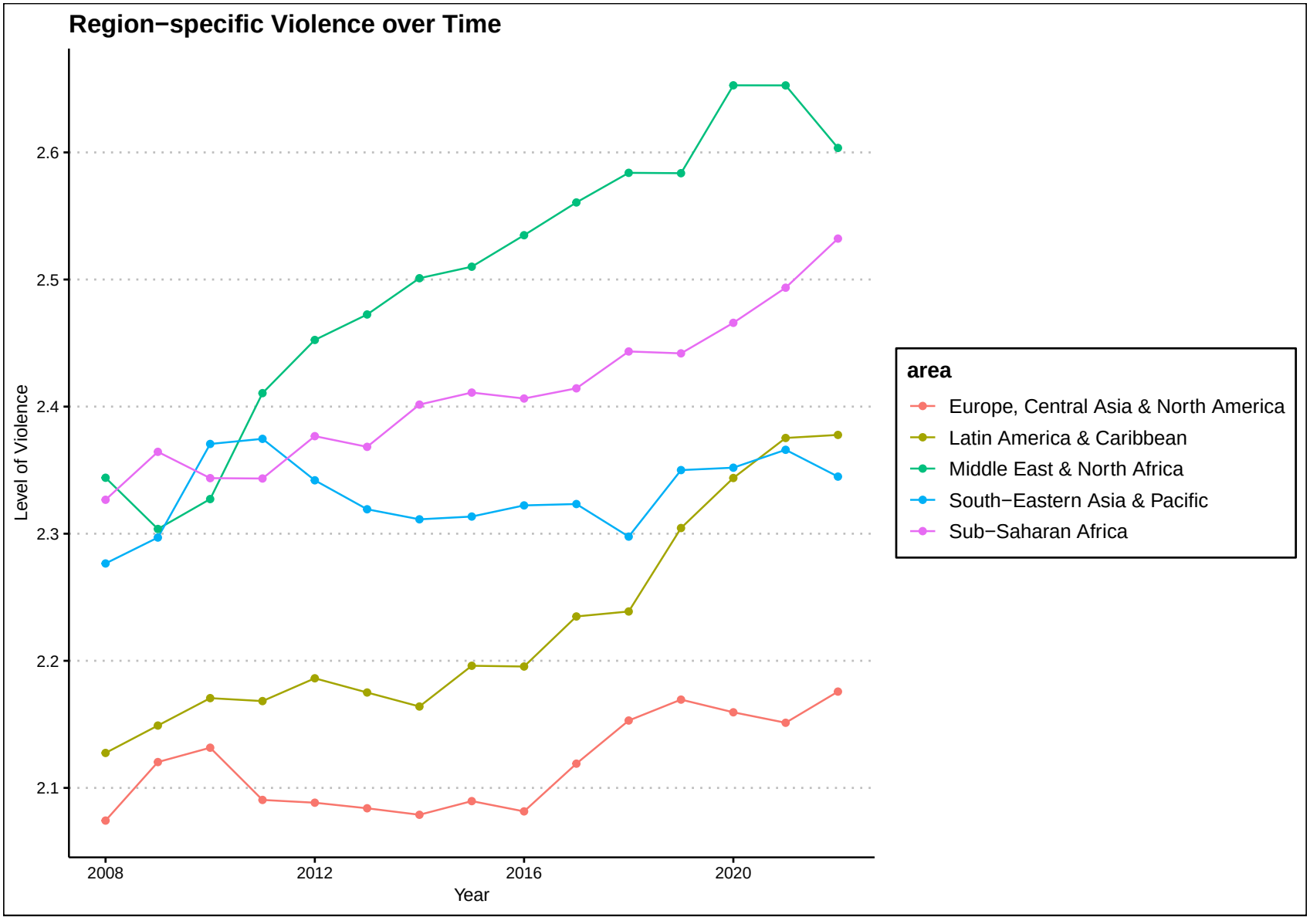
```
data_violence_region = data |>
      group_by(Year,area) |>
      summarize(GPI=weighted.mean(x=GPI,w=pop,na.rm = FALSE))
```

2. Produce the plot

```
plot1 = ggplot(data=data_violence_region, # data input
     mapping=aes(x=Year,y=GPI,color=area))+ # relationships
     geom_line()+ # add a line for each country
     geom_point()+ # add points
     theme_clean()+ # specify a theme
     xlab('Year')+ #label for title of x-axis
     ylab('Level of Violence') + # label for title of y-axis
     ggtitle('Region-specific Violence over Time')  # title
```

Question (1) cont.

```
plot1
```



Question (2)

Create a similar visual aid for life expectancy

1. Calculate region-specific life expectancies

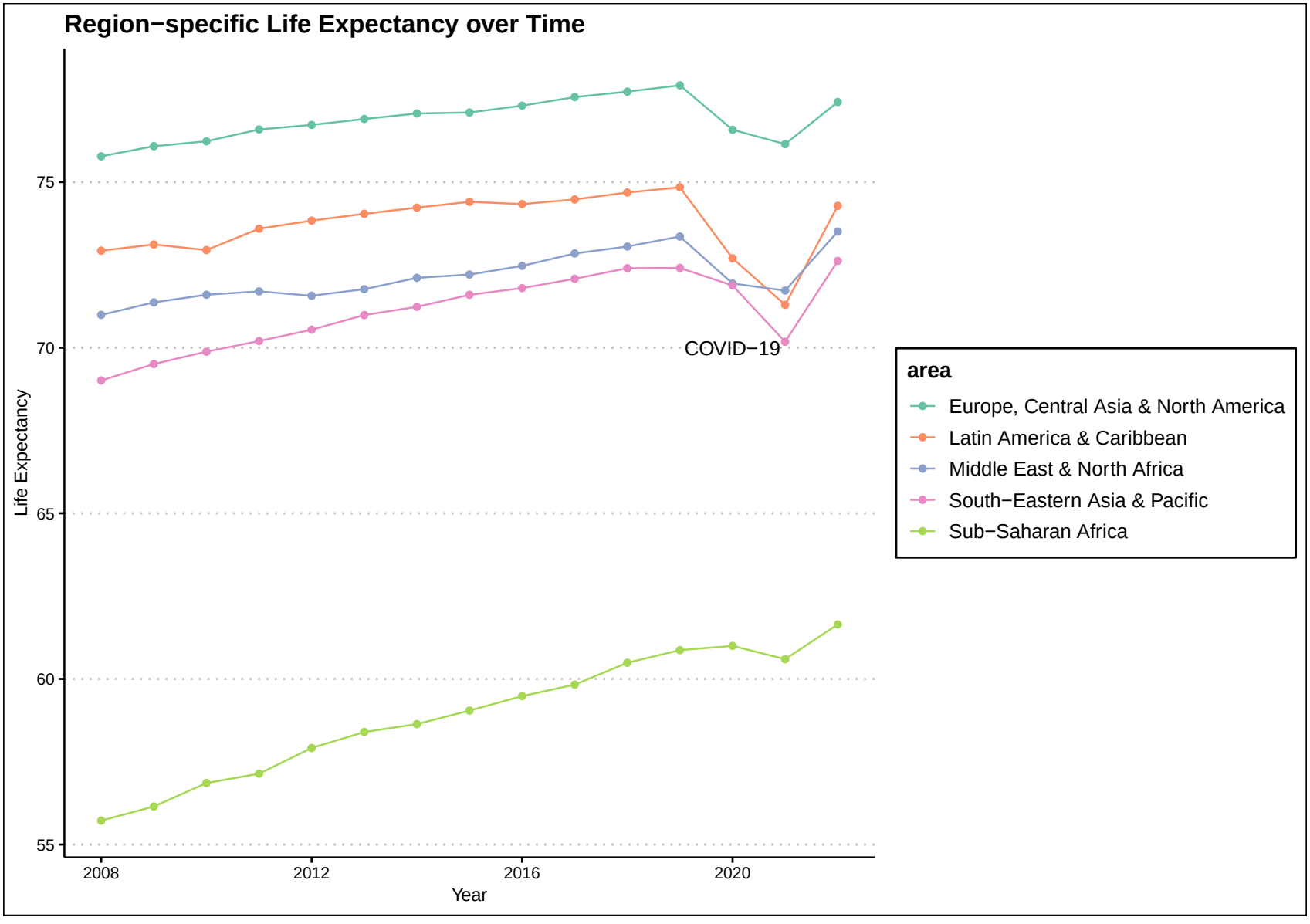
```
data_life_exp_region = data |>
      group_by(Year,area) |>
      summarize(e0=weighted.mean(x=e0,w=pop,na.rm = FALSE))
```

2. Produce the plot

```
plot2 = ggplot(data=data_life_exp_region, # data input
     mapping=aes(x=Year,y=e0,color=area))+ # relationships
     geom_line()+ # add a line for each country
     geom_point()+ # add points
     annotate("text",x=2020,y=70,label="COVID-19")+
     theme_clean()+ # specify a theme
     xlab('Year')+ #label for title of x-axis
     ylab('Life Expectancy') + # label for title of y-axis
     scale_color_brewer(palette = "Set2")+
     ggtitle('Region-specific Life Expectancy over Time')
     # title
```

Question (2) cont.

```
plot2
```



Question (3)

Produce two scatter plots to display the relationship between violence and life expectancy in 2008 and 2022. Add also a fitting line.

Select records for years 2008 and 2022

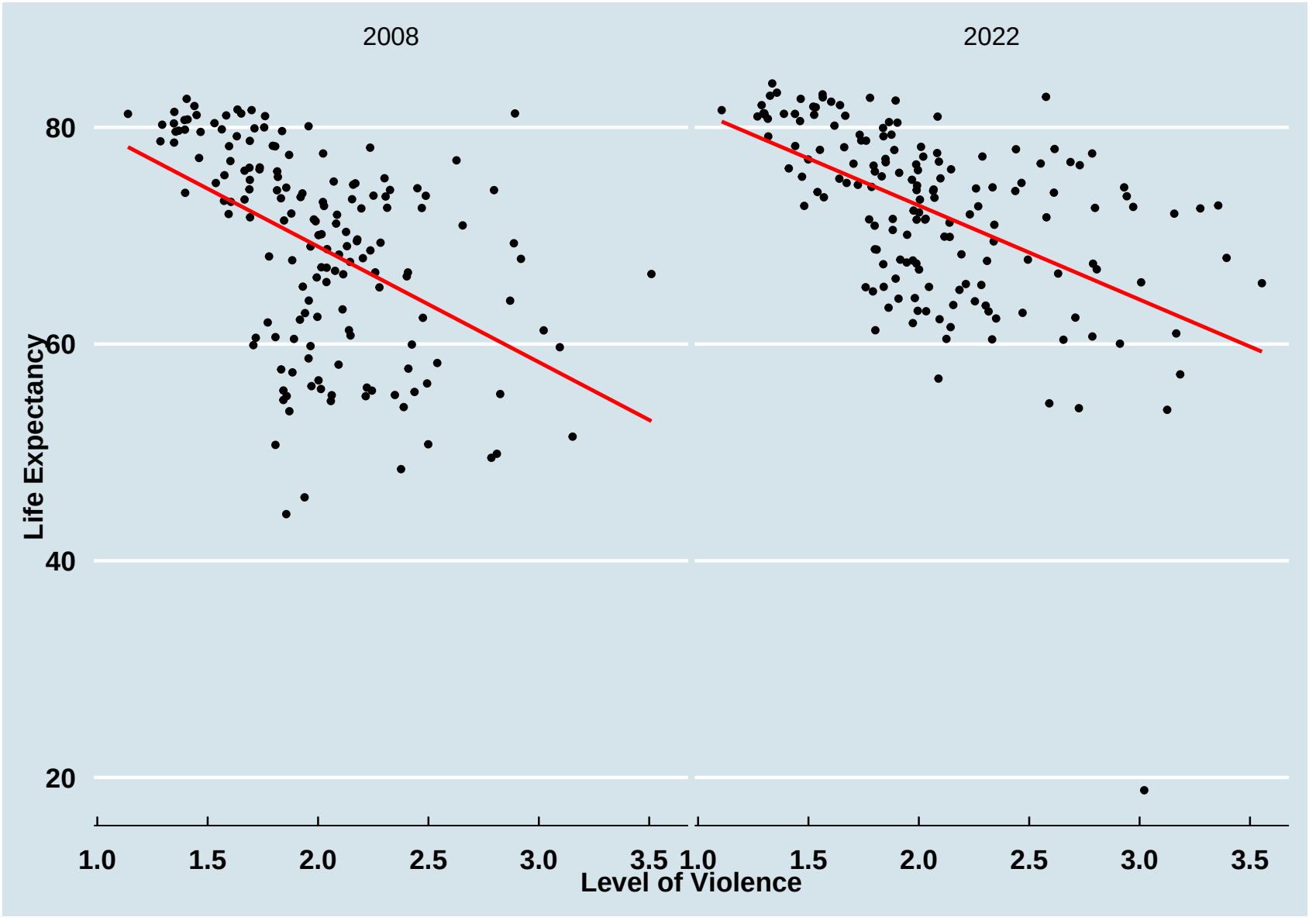
```
data_scatter = data |>
filter(Year %in% c(2008,2022))
```

Select records for years 2008 and 2022

```
plot3 = ggplot(data=data_scatter, # data input
mapping=aes(x=GPI,y=e0))+ # relationships
geom_point()+
xlab('Level of Violence')+
ylab('Life Expectancy')+
theme_economist()+
geom_smooth(method = "lm", se = FALSE,color='red')+ # regression line
facet_wrap(~Year) + # seperate by year
theme(axis.text.y = element_text(size=15,face="bold"), # font of y-axis text
      axis.title.y = element_text(size=15,face="bold"), # font of y-axis title
      axis.text.x = element_text(size=15,face="bold"),
      axis.title.x = element_text(size=15,face="bold")) # font of x-axis title
```

Question 3. (cont.)

```
plot3
```



Question 4

Display the distribution of life expectancy in the 20 most violent countries and in the 20 most peaceful countries in 2022

Let's create the data sets Most violent countries

```
data_most_violence = data |>
       filter(Year==2022) |>
      slice_max(GPI,n=20) |>
      mutate(label='Most Violent')
```

Most peaceful countries

```
data_most_peaceful = data |>
filter(Year==2022) |>
slice_min(GPI,n=20) |>
mutate(label='Most Peaceful')
```

combine the two data sets by row

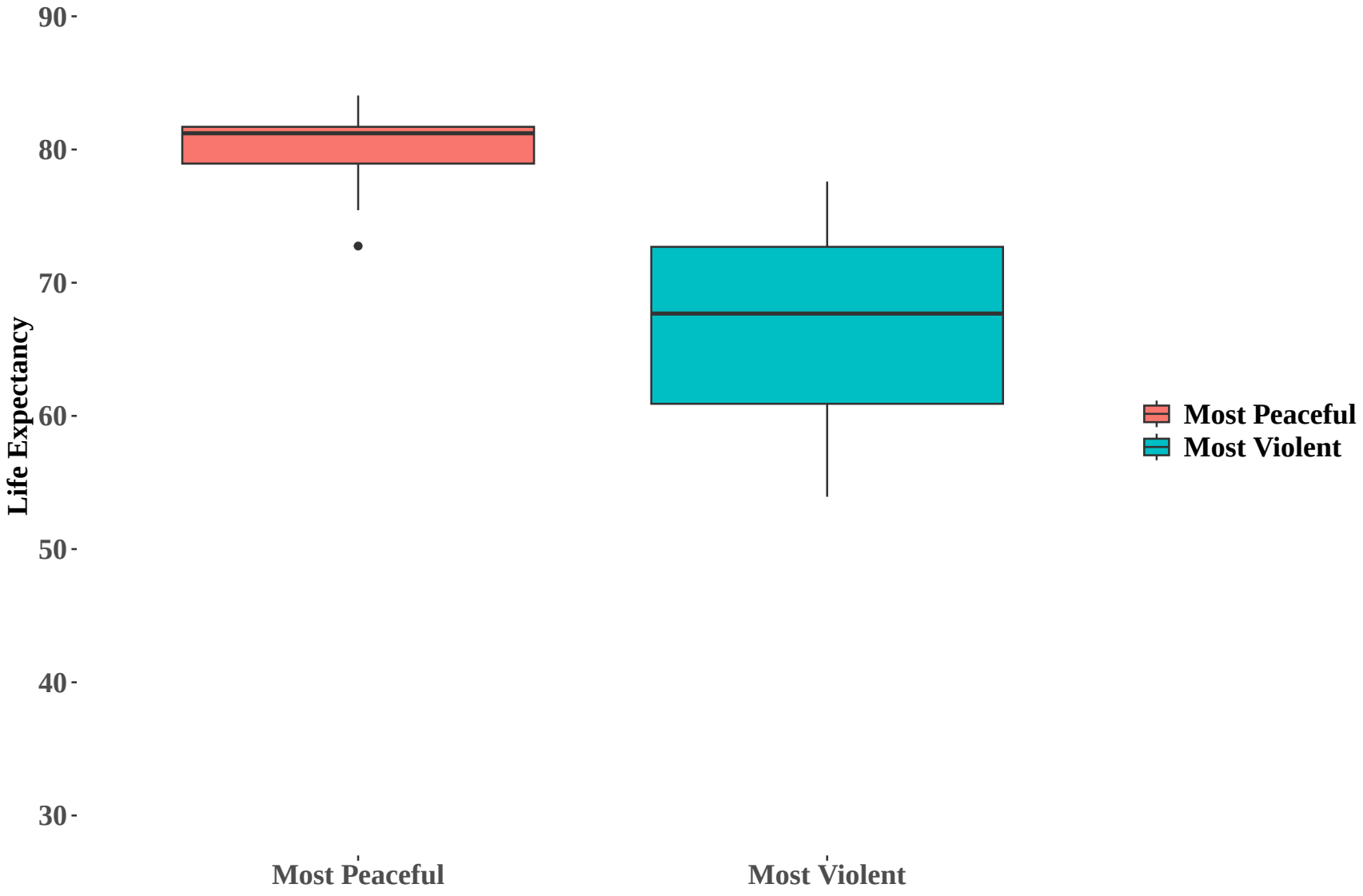
```
data_plot=rbind(data_most_violence,data_most_peaceful)
```

Question 4 (cont.)

```
plot4=ggplot(data=data_plot,
       aes(x=label,y=e0,fill=label))+
       geom_boxplot()+
    coord_cartesian(ylim=c(30,90))+
     scale_y_continuous(breaks=seq(30,90,10),
                        labels=seq(30,90,10))+
     theme_tufte()+
     scale_fill_discrete(name='')+
     xlab('')+
     ylab('Life Expectancy')+
     theme(axis.text.y = element_text(size=15,face="bold"), # font of y-axis te
      legend.text = element_text(size=15,face="bold"),
      axis.title.y = element_text(size=15,face="bold"), # font of y-axis title
      axis.text.x = element_text(size=15,face="bold"),
      axis.title.x = element_text(size=15,face="bold")) # font of x-axis title
```

Question 4 (cont.)

```
plot4
```



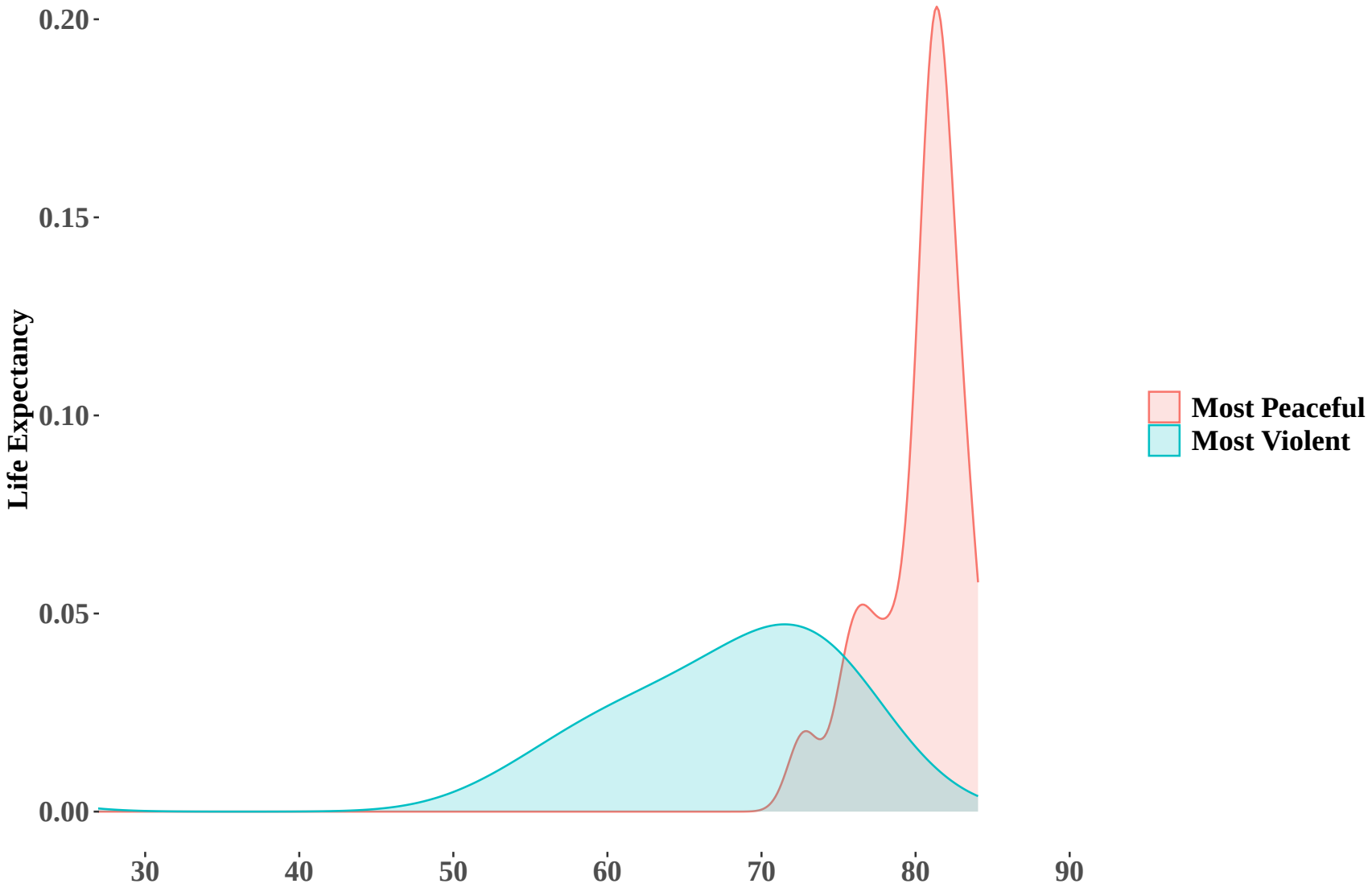
Question 5

Perform the same task using a different visual aid.

```
plot5 = ggplot(data=data_plot,
       aes(x=e0,fill=label,color=label))+
     geom_density(alpha = 0.2, na.rm = TRUE) +
     theme_tufte()+
     scale_fill_discrete(name='')+
     scale_color_discrete(name='')+
     coord_cartesian(xlim=c(30,90))+
     scale_x_continuous(breaks=seq(30,90,10),
                        labels=seq(30,90,10))+
     xlab('')+
     ylab('Life Expectancy')+
     theme(axis.text.y = element_text(size=15,face="bold"), # font of y-axis te
      legend.text = element_text(size=15,face="bold"),
      axis.title.y = element_text(size=15,face="bold"), # font of y-axis title
      axis.text.x = element_text(size=15,face="bold"),
      axis.title.x = element_text(size=15,face="bold")) # font of x-axis title
```

Question 5 (cont.)

```
plot5
```



Quarto

Quarto enables you to weave together content and executable code into a finished presentation. To learn more about Quarto presentations see https://quarto.org/docs/presentations/.

Bullets

When you click the **Render** button a document will be generated that includes:

- Content authored with markdown
- Output from executable code

Code

When you click the **Render** button a presentation will be generated that includes both content and the output of embedded code. You can embed code like this:

```
gpi_data = read.table('Data/gpi_data.txt',header=T)
life_exp_data = read.table('Data/life_exp_data.txt',header=T)
x=gpi_data |>
  left_join(life_exp_data,by=c('Year','iso3'))
```